Enclosure (1) to JWR-04-038

JWR-04-010
June 2004
JNC03

The Referent Study Final Report

by: D. K. Pace

The Referent Study Final Report*

June 2004

Study Team (participants from China, France, UK, and US)
Study Lead: Dale Pace (Johns Hopkins University Applied Physics Laboratory, JHU/APL)
Richard Bernstein (JHU/APL)
Zhang Bing (Harbin Institute of Technology, China)
Mark Dumble (Aegis Simulation Technologies UK Ltd, United Kingdom)
Daniel Girardot (Centre d'Analyse de Défense, France)
COL Kevin Greaney (Missile Defense Agency, MDA)
Robert Hamber (Naval Facilities Engineering Command, NAVFAC)
Hanae Hara (Space and Naval Warfare Systems Command, SPAWAR)
Scott Harmon (Zetetix)
Sam Johnson (USAF Modeling and Simulation)
Mike Leite (SETA support to DMSO, SAIC)
Mike Metz (IMC)
Stuart Randlett (SAIC)
Bob Senko (DMSO Consultant)
Susan Solick (Army Training and Doctrine Analysis Center, TRAC)
Marcy Stutzman (Northrop Grumman)
James Wallace (Army Modeling & Simulation Office, Alion)
Scott Weidman (National Academy of Sciences)
Simone Youngblood (DMSO VV&A Technical Director, JHU/APL)

Objectives of the Referent Study

1. To articulate a definition and connotation for *referent* which is suitable for VV&A applied to the full spectrum of models and simulations (M&S).
2. To identify and articulate desired *referent* characteristics.
3. To develop guidelines (based upon current state of the M&S art and best VV&A praxis) for selecting and describing a *referent*.
4. To develop guidelines (based upon current state of the M&S art and best VV&A praxis) for *referent* use in validation and accreditation assessments.
5. To identify research needs for advancing *referent* identification/selection, articulation, and use in validation and accreditation assessments.

*This study was chartered by the Verification, Validation, and Accreditation (VV&A) Technical Director of the U.S. Defense Modeling and Simulation Office (DMSO). Views of this report are those of the Study Team and should not be construed to represent views of DMSO or of any other organization or agency, public or private. The report views are primarily those of the Study Lead who was its principal author and who redacted contributions from other members of the Study Team. It should not be presumed that every member of the Study Team subscribes to everything in this report, although it is believed that the report fairly represents views of the Study Team.

Table of Contents

Table of Contents (continued)

List of Figures

Acknowledgments

This study has both breadth and depth because of the varied background and perspectives of the individuals who comprise the study team. Appreciation is hereby expressed both to them and to their organizations, whose support made possible their participation.

Executive Summary

The importance of reliable and complete information about what a model or simulation (M&S) is to represent and about what should be the standard that M&S results will be compared with has long been recognized. In spite of such recognition, the modeling and simulation (M&S) literature does not address extensively how to identify, select, or describe the information about what a M&S is to represent for use in validation and accreditation assessment. This report calls that information the *referent*.

Lack of extensive treatment of the referent in M&S literature has hindered development of best practices and contributed to inconsistent performance by M&S communities. The Verification, Validation, and Accreditation (VV&A) Technical Director of the U.S. Defense Modeling and Simulation Office (DMSO) chartered this study of the referent for validation and accreditation assessments. While this study has not produced the final word on the referent, it is a substantial contribution to the M&S literature. Nineteen people from four countries (China, France, United Kingdom, and United States) contributed to or reviewed the study.

Referent Study Objectives

- To articulate a definition and connotation for *referent* which is suitable for VV&A applied to the full spectrum of M&S.
- To identify and articulate desired *referent* characteristics.
- To develop guidelines (based upon current state of the M&S art and best VV&A praxis) for selecting and describing a *referent*, and as needed to make such guidelines a function of M&S characteristics and applications.
- To develop guidelines (based upon current state of the M&S art and best VV&A praxis) for *referent* use in validation and accreditation assessments, and to include suggestions for what can be done when the referent is inadequate, poorly identified, or unknown.
- To identify research needs for advancing *referent* identification/selection, articulation, and use in validation and accreditation assessments.

What Is the Referent?

A variety of definitions exist for referent, but none of them quite meet the needs of the referent to support M&S validation and accreditation assessments. For that purpose, we define referent as:

> *The referent is the best or most appropriate codified body of information available that describes characteristics and behavior of the reality represented in the simulation from the perspective of validation assessment for intended use of the simulation.*

The referent consists of information (data, theory, calculations, expert opinion, or combinations of these) that is systematic and organized (preferably from an authoritative source). The reality represented includes actors, systems and entities interacting with other actors, systems and entities through processes in one or more environments. In general, the most appropriate information is the least expensive that has adequate fidelity (accuracy, scope, resolution, context) to serve as the basis for comparison in validation and accreditation assessments for intended M&S use.

When Should the Referent Be Specified?

A referent must have adequate fidelity to support validation assessment. Referents may be selected by direction, convenience, economics, or proxy (using information about something similar). The referent may be specified at any time during the M&S life cycle, but the best place to specify the referent is in the simulation conceptual model.

How Should the Referent Be Described?

While the information content of the referent is dependent upon the kind of M&S and its application domain as well as intended use, identification and specification of the referent should be definite and unambiguous in all cases. The referent description should specify the referent context, its domain coverage, and pertinent actor/system/entity/process/environment attributes. Parameter uncertainties should be quantified and how information from various sources is combined to form the referent should be explained.

Referent Use in Validation and Accreditation Assessments

In data-rich environments, it is best if there is statistical independence between information used in M&S development (or operation) and the information used as the referent for validation and accreditation assessment. This allows the most robust assessment of M&S predictive capabilities.

In data-poor environments, reliance upon theoretical information and dependencies between the referent and M&S development information limit capability for reliable judgment about M&S predictions.

If the referent is inadequate (e.g., the referent has unresolved contradictory information), it may be impossible to perform a meaningful validation assessment of the M&S for its intended use.

Referent Research Needs

There are three primary referent research needs: 1) how to consolidate referent information so that it is coherent and uncertainties are appropriately quantified; 2) how to specify and describe dynamic referents; and 3) how to document referent use in M&S validation and accreditation.

1. Introduction

1.1 Background

The importance of reliable and complete information about what a model or simulation (m/s) is to represent and about the standard that M&S results will be compared with has long been recognized. It is a truism that representational fidelity is restricted by the completeness and detail in the definition of what is to be modeled. Restrictions imposed on modeling and simulation (M&S) utility by limitations in such information has led some to develop methods such as exploratory analysis to compensate for such limitations [e.g., Bigelow and Davis, 2003]. This problem has led others to an emphasis on improving quantification of uncertainties in simulation and experimental results to reduce those limitations [e.g., Oberkampf et al, 2000]. In spite of recognition of the importance of information about what an M&S represents and many different approaches to overcome limitations in such, none of the M&S communities have a) formalized terms for such information, b) provided general guidance for how to select such information or describe it, c) created paradigms that show how such information varies with subject or application domain, or d) developed widely accepted approaches for how to use that information. The M&S literature does not address substantively how to identify, select, or describe the information about what an M&S is to represent for use in validation and accreditation assessment. This report calls that information the *referent* (formally defined later). Some consider emphasis upon formal referent definition and processes a new notion that opens a wide range of research studies. Only a few papers, such as Girardot and Jacquart [2002] drawing upon a United Kingdom and France common study, have grappled with this issue.

This lack of extensive and explicit treatment of the referent in the literature has hindered development of best practices and contributed to inconsistent and sometimes inept performance by M&S communities. This report was stimulated by this situation, and provides a substantial contribution to the M&S literature about the referent[1]. We do not pretend that we have produced the final word on this subject, but we believe that we have made a substantial contribution to ideas about the referent and hope that this report will provide the impetus needed for the M&S world to bring understanding of the referent to maturity so that M&S will benefit.

1.1.1 Why A Referent Is Needed

For an M&S to be used with confidence, it must undergo examination for correctness, fitness and utility, which are measured against acceptability criteria that include metrics for accuracy, resolution, tolerance, etc. Information selected to be the basis for those criteria is the referent. The referent is the standard against which to calculate errors of a simulation (i.e., differences between simulated values of properties and values actually observed in the simuland). Measures of correctness, fitness and utility are the basis for validating and accrediting the m/s.

---

[1] A variety of referents exist in M&S constructs [e.g., Gross, 1999]. Usually a descriptor indicates focus or domain of the referent. This report deals with only one kind of M&S referent; the referent used in validation and accreditation assessments. For convenience, no adjective is attached to the term, though we frequently state this focus so those who only read parts of this report will not be confused or misunderstand. This restriction provides content for the report. Many of the ideas in this report should be useful for other kinds of referents too.

1.2 Objectives and Scope of Study

1.2.1 Study Sponsorship

This study was instigated by Ms. Simone M. Youngblood, Technical Director for Verification, Validation, and Accreditation (VV&A) of the Defense Modeling and Simulation Office (DMSO) early in 2004. Most of the Referent Study participants were drawn from two overlapping groups: 1) the Department of Defense (DoD) VV&A Technical Working Group (TWG), and 2) those involved with the VV&A Forum of the Simulation Interoperability Workshop (SIW) of the Simulation Interoperability Standards Organization (SISO). Although the Referent Study was performed under the auspices of the DMSO VV&A Technical Director, views of this report are those of the Study Team and should not be construed to represent views of DMSO or of any other organization or agency, public or private. More precisely, the report views are those of the Study Lead (Dr. Dale K. Pace) who was its principal author and who redacted contributions from other members of the Study Team. It should not be presumed that every member of the Study Team subscribes to everything in this report, although it is believed that the report fairly represents views of the Study Team.

1.2.2 Study Objectives

The Referent Study has five objectives:

- To articulate a definition and connotation for *referent* which is suitable for VV&A applied to the full spectrum of models and simulations (M&S).

- To identify and articulate desired *referent* characteristics.

- To develop guidelines (based upon current state of the M&S art and best VV&A praxis) for selecting and describing a *referent*, and as needed to make such guidelines a function of M&S characteristics and applications.

- To develop guidelines (based upon current state of the M&S art and best VV&A praxis) for *referent* use in validation and accreditation assessments, and to include suggestions for what can be done when the referent is inadequate, poorly identified, or unknown.

- To identify research needs for advancing *referent* identification/selection, articulation, and use in validation and accreditation assessments.

All of these objectives are addressed in this report, although some objectives are more fully addressed than other objectives.

1.2.3 Study Scope

The scope of the Referent Study is limited by its short duration, 3-4 months. DMSO funding specifically for this project only provided the Study Lead. Appreciation for the importance of the subject addressed made possible the voluntary participation of the other members of the Study Team. The Referent Study will have four formal products; two of which were produced in March and April, the third is this report, and the fourth comes later:

1. A briefing to the DoD VV&A Technical Working Group (TWG) at its March 2004 meeting (see Appendix B).

2. A briefing to the VV&A Forum at the Spring 2004 Simulation Interoperability Workshop (SIW) in April on plans for the Referent Study (see Appendix C).

3. A final report of the Referent Study (this document), which will either be posted on the DMSO VV&A web page or linked to it, so that insights from the study can be widely available.

4. A paper based upon the final report of the Referent Study for the 2004 European Simulation Interoperability Workshop (EuroSIW) in Edinburgh (Scotland) in June.

Other uses of ideas from the Referent Study (such as papers at other conferences and workshops, materials in future updates to the *DoD M&S VV&A Recommended Practices Guide [RPG]*, etc.) are expected, but such will not be "formal" products from the Referent Study. An important product that we expect from this report and the EuroSIW paper is a substantial increase in referent discussion, debate, and suggestions in the M&S literature.

1.2.3 Report Organization

The Referent Study report has five main components:

- Referent Definition and Connotation
- How to Identify Possible Referents and Select an Appropriate Referent
- How to Specify/Describe the Referent
- Referent Use in M&S Development and in Validation and Accreditation Assessments
- Referent Research Needs

The report ends with a Conclusion and References. Appendices contain the two briefings about plans for this study plus acronyms and definitions of pertinent terms.

1.2.5 Study Participants

Primary participants in the Referent Study are identified below. They come from China, France, the United Kingdom (UK) and the United States (US), and bring many different M&S perspectives, which help this report to have both the breadth and depth it needs to be most useful.

Individual attribution for sections is not provided since all sections of the report were reviewed by more than one person, and some sections represent comments from several or many members of the Study Team.

Study Lead: Dale Pace (Johns Hopkins University Applied Physics Laboratory, JHU/APL)
Richard Bernstein (JHU/APL)
Zhang Bing (Harbin Institute of Technology, China)
Mark Dumble (Aegis Simulation Technologies UK Ltd, United Kingdom)
Daniel Girardot (Centre d'Analyse de Défense, France)
COL Kevin Greaney (Missile Defense Agency, MDA)
Robert Hamber (Naval Facilities Engineering Command, NAVFAC)
Hanae Hara (Space and Naval Warfare Systems Command, SPAWAR)
Scott Harmon (Zetetix)
Sam Johnson (USAF Modeling and Simulation)
Mike Leite (SETA support to DMSO, SAIC)
Mike Metz (IMC)
Stuart Randlett (SAIC)
Bob Senko (DMSO Consultant)
Susan Solick (Army Training and Doctrine Analysis Center, TRAC)
Marcy Stutzman (Northrop Grumman)
James Wallace (Army Modeling & Simulation Office, Alion)
Scott Weidman (National Academy of Sciences)
Simone Youngblood (DMSO VV&A Technical Director, JHU/APL)

2. Referent Definition and Connotation

2.1 Existing Definitions

A variety of definitions exist for *referent*. Some definitions are shown below; the definitions are in **bold**. Their sources are indicated.

- The glossary of the DoD M&S VV&A *Recommended Practices Guide (RPG)* contains two definitions for *referent*:
  - **A codified body of knowledge about a thing being simulated.** The result of Simulation Interoperability Workshop (SIW) Fidelity Implementation Study Group (ISG) discussions and comments from March 1998 to December 1998.
  - 2. **Something referenced or singled out for attention, a designated object, real or imaginary or any class of such objects.** From Heylighen and Webster's II, New College Dictionary, both cited below.
- **Something pointed to or singled out for attention, a designated object, real or imaginary or any class of such objects.** F. Heylighen, *Web Dictionary of Cybernetics and Systems*, http://pespmcl.vub.ac.be/ASC/indexASC.html, nd.
- **Something referred to.** Houghton Mifflin Co., *Webster's II, New College Dictionary*, 1995.
- **Someone that is referred to or consulted; a word or term that refers to another; (in logic) the term from which a relationship proceeds; that which is denoted or named by an expression or a statement.** G. & C. Merriam Company, *Webster's Third New International Dictionary (Unabridged)*, 1971.
- The following definitions for *referent* may be found on the web:
  An assumed zero value of a quantity relative to which magnitudes of the quality are measured, or a structure having this zero value of the quantity; e.g., a voltage measured relative to the ground as a referent. vesuvius.jsc.nasa.gov/er/seh/r.html
  The referent of a metalevel is that term it is metalevel of. The referent of a task is the result value of that task. www.iiia.csic.es/~enric/noos/Overview/node32.html
  1. Something that refers, especially a linguistic item in its capacity of referring to a meaning; 2. Something referred to tsolutions.net/help/glossaryM-Z.htm
  The object or event (real or imagined) to which a word (symbol) refers. www.casad.org/English113/Writing/glo.htm
  Something referred to; the object of a reference www.cogsci.princeton.edu/cgi-bin/webwn
  The first term in a proposition; the term to which other terms relate www.cogsci.princeton.edu/cgi-bin/webwn
  Something that refers; a term that refers to another term www.cogsci.princeton.edu/cgi-bin/webwn

Each of the definitions above is correct and valid, but none of them may be the best definition for a referent in the context of M&S VV&A. The definition below is proposed for this study because it seems more appropriate for the objectives of this study. This definition may also be very useful for M&S VV&A in general. None of the definitions above explicitly requires the referent to be the "best" or "most appropriate" information available. The "codified" aspect of the referent from the *RPG* and the Fidelity ISG brings an implied connotation of systematic organization to the knowledge or information that is helpful. The definition proposed below is followed by a discussion of its connotation.

This study is focused on the role of the referent in validation and accreditation assessments. We do not address how information used to develop a model or simulation is identified or described. Nor do we address how one selects data that is appropriate to use as simulation inputs, whether hardwired in the simulation or fed into the simulation as needed to run it. We restrict our study to the referent associated with M&S validation and accreditation assessments; however, this makes us include issues related to statistical independence of information used for simulation development and information used as the referent in validation and accreditation assessments when simulation results are to be used to predict performance or behavior.

2.2 Proposed Definition and Rationale

Referent: *The referent is the best or most appropriate codified body of information available that describes characteristics and behavior of the reality represented in the simulation from the perspective of validation assessment for intended use of the simulation.*

Various words and phrases from this definition are discussed below to ensure that it is fully understood. Rationale for why it is a better definition of referent for the purposes of this study than existing ones in the community, as presented above, is also presented.

The **information** used as the referent may consist of:

- "data" (observations of the simuland, either under controlled circumstances as in tests and experiments, or under natural circumstances),
- theories as expressed in algorithms that describe characteristics, behaviors and relationships (preferably theories validated against observations of the simuland),
- simulation results (preferably from simulations that have been objectively and quantitatively validated),
- expert human knowledge, or
- combinations of these.

In cases where explicit observations and theories do not provide a comprehensive and sufficiently reliable description of the reality represented in the M&S, information from subject matter experts (SMEs) may have to serve as the referent or part of it. Such SME information may not be explicitly articulated and systematically organized since it is a form of qualitative assessment, and the basis for qualitative assessments usually are not as explicitly described as in quantitative assessments. However, it is recognized that M&S communities that deal mostly with physics-based models that may in principle be compared against quantitative data tend to be less willing to accept SME information and other qualitative assessments as part of the referent than are other M&S communities. The computational science and engineering community, which has major verification and validation (V&V) concerns in applications of computational fluid dynamics (CFD) and computational solid mechanics simulations, is an example of that kind of M&S community.

The information used as a referent is a **codified body of information**, with that body of knowledge drawn from one or more information sources. Codify has several connotations. It implies system and organization, both useful aspects of information to be used as a referent for M&S validation. Codified often implies authority and is frequently associated with laws. Wherever appropriate and possible, the referent should be information drawn from credible authoritative data sources.

The definition above deals with the "**reality represented in the simulation**." Typically that *reality* includes actors/systems/entities interacting with other actors/systems/entities by various processes through or in one or more environments. The referent pertains to all of these: actors/systems/entities, processes, and environments. Connotations associated with these may vary by M&S type and by the kind of application. For example, the simulation operator would not be part of the referent for a batch-run constructive simulation, but the simulation operator might need to be part of the referent in an interactive simulation (such as a game or war game).

There are three challenges to creating a referent defined as the **best or most appropriate** information available that describes **characteristics and behavior** of the reality represented in the simulation **from the perspective of validation assessment for intended use of the simulation.**

- Development and organization of the **best** information possible may be too costly and/or take too long for schedule and resource constraints of a particular M&S application. This leads to use of an *adequate referent*; one that is not the "best" information possible but one that is adequate as the basis for validation assessment within the context of the M&S intended use. Such an adequate referent may be the "best" information possible within the time and resource constraints for a particular M&S application.

  The idea of an "adequate referent" increases the importance of exact and precise specification of intended M&S use. A poorly specified intended use increases the risk of decision error, i.e., the risk either that a simulation will be judged acceptable (valid) for the intended use when it is in fact not acceptable, or that an acceptable M&S will be judged unacceptable. Consequences of such mistakes depend upon the impact of the M&S application. There may be little consequence from such a decision error for a simulation that only produces background information about a subject, but catastrophe could result from such a decision error about a simulation used as part of a real-time decision aid for a safety-critical system. The ultimate authority for the acceptability of a referent (i.e., whether the referent is adequate for intended use) is the accrediting activity (or accreditation authority), the person or organization responsible for the decision that the M&S is appropriate for the intended use. This report develops rationale for identification, selection, and description of a referent by the V&V agent (or others responsible for such) and a basis for its acceptance by the accrediting activity.

- The **most appropriate** information has adequate fidelity[2] (accuracy, scope, resolution, context) to serve as the basis for comparison in validation assessment of the M&S for its intended use. It is a truism that one cannot demonstrate greater fidelity for simulation results than the fidelity of the referent to which those results are compared. Consequently, an M&S validation assessment requires that the referent have greater fidelity (i.e., more closely represent the reality addressed by the m/s) than the required simulation results. Vague specification of M&S intended use can make it difficult to determine required fidelity for an acceptable referent. This leads to increased risk that the M&S will be erroneously considered valid for the intended use.

  As a general rule, the most appropriate information to serve as the referent is the least expensive set of information that has adequate fidelity to support validation assessment for M&S intended use. For example, if simulation results are to provide ballpark estimates of performance as background information, inexpensive SME judgment may be an acceptable referent with adequate fidelity for a validation assessment. Spending time and effort collecting more precise (increased fidelity) data and processing the data into an acceptable referent format could be considered a waste of resources, or "gold plating."

- Because the referent **describes characteristics and behavior of the reality represented in the simulation,** specification of the referent is particularly difficult for those M&S which generate new knowledge or additional knowledge about the reality represented in the M&S (vice the M&S simply being a reliable representation of that reality). This could be the case with a simulation demonstrating emergent behavior. This also could be the case if the simulation employs some form of judgment in uncertain conditions (such as representation of human decision making) or if the simulation portrays a future reality that is difficult to predict (such as social structure, military posture, etc.). There are two implications here for referent specification. One is that iteration of the referent specification may be necessary as new knowledge is generated. The hazards of such should be obvious. The other is that M&S intended use with such uncertainties in the reality represented implies that the referent specification will have low fidelity. It is not the length of time projected into the future that is the issue; astrophysical simulations project conditions billions of years in the future with high fidelity. It is the uncertainty in the representation that determines the fidelity needed in the referent specification.

2.3 Connotations

We do not attempt to cover all possible connotations of referents in the study; in the portion of the study that addresses how referents should be described, we examine ten varieties of models

[2] This report uses a general breakdown of fidelity (accuracy, scope, resolution, context) instead of the more elaborate construct proposed by the Simulation Interoperability Workshop (SIW) Fidelity Implementation Study Group [Gross, 1999] or its follow-on Fidelity Experimentation Implementation Study Group [Roza et al, 2000]. The reason for this is simple. The more elaborate construct has not been formally adopted by the Simulation Interoperability Standards Organization (SISO), to which SIW belongs, nor any other major M&S body. A general description is adequate for this study. Our description is compatible with the way fidelity is treated in the Institute Electrical and Electronics Engineers (IEEE) Standard 1516.3 of *Recommended Practice for High Level Architecture (HLA) Federation Development and Execution Process (FEDEP).*

and simulations specifically to give the study reasonable breadth and depth in terms of connotation variation for referents.

Some contend that the referent should be independent of M&S variety or application. We believe that while the definition we propose can apply to all varieties of M&S and their full range of potential applications, how one identifies the information appropriate for a referent and selects an appropriate set of information as well as how one describes the referent will vary at times with M&S type and application. Material presented in Section 4 illustrates this.

Many M&S communities, especially those in which models often have heuristic factors that are adjusted to make simulation results fit experimental data (as is the case, for example, in computational solid mechanics) stress the importance of separate data for results used in such model calibration from the data used in M&S validation assessments. The problem is similar to that encountered in clinical trials in the health field, in which randomization of patient involvement in a clinical trial (with its associated unpredictability) is used to "protect against the unpredictability of the extent of bias in the results of non-randomized clinical trials" (Kunz and Oxman, 1998). Many do not appreciate the issues associated with using the same information for M&S validation assessment as used for M&S development.

For simplicity, we shall use validation assessment to encompass all referent aspects of both validation and accreditation assessment. It is important to appreciate the difference between validation assessments and accreditation assessments. The validation assessment determines whether the M&S has adequate fidelity to support intended uses, and perhaps can even quantify the likelihood that the M&S has the required fidelity with a statement like "we have 90% confidence that M&S results will differ from the referent by less than 5% for any conditions within the specified application region." Such a quantitative statement is the most precise kind of validation assessment, and depends upon a very well defined intended use, an explicitly well defined referent, and a precise measurement capability for the m/s. Few validation assessments have such quantitative precision.

Accreditation assessment uses the validation assessment, but supplements it with risk, programmatic and other considerations to determine if the M&S is acceptable for use in a particular situation. *Whenever accreditation assessment is mentioned in this report, we are only concerned with its validation component.* This makes it possible for us to use the term validation assessment to refer to either or both validation and accreditation assessments. It is important to remember that validation assessment is always in the context of intended M&S use, and the accreditation activity (or accreditation authority) is the one who determines the acceptability criteria and in essence is the one to approve the referent.

3. How to Identify Possible Referents and Select an Appropriate Referent

M&S validation and accreditation assessments may be done in a formal process, as is often the case with a major M&S application, or they may be done very informally as may be the case in a minor M&S application. In either case, the same kind of processes are involved in deciding that an M&S is appropriate for a particular application. This section is concerned with two topics germane to these processes. First how to identify possible referents for M&S validation assessment; and, second, how to select an appropriate referent when there is more than one possible referent.

3.1 Referent Identification

Two factors make identification of a M&S validation referent non-trivial: a) M&S requirements, and b) variations in characteristics of the reality to be represented in the m/s. Requirements identify what is to be represented in the M&S (systems, entities, environments, etc.) and what is the intended use of the M&S (which has implications for the fidelity of representations and how they can be manipulated). A particular application may be driven by a subset of M&S requirements. That subset is called the acceptability criteria. Figure 3.1-1 illustrates how these factors relate to one another in an accreditation assessment.

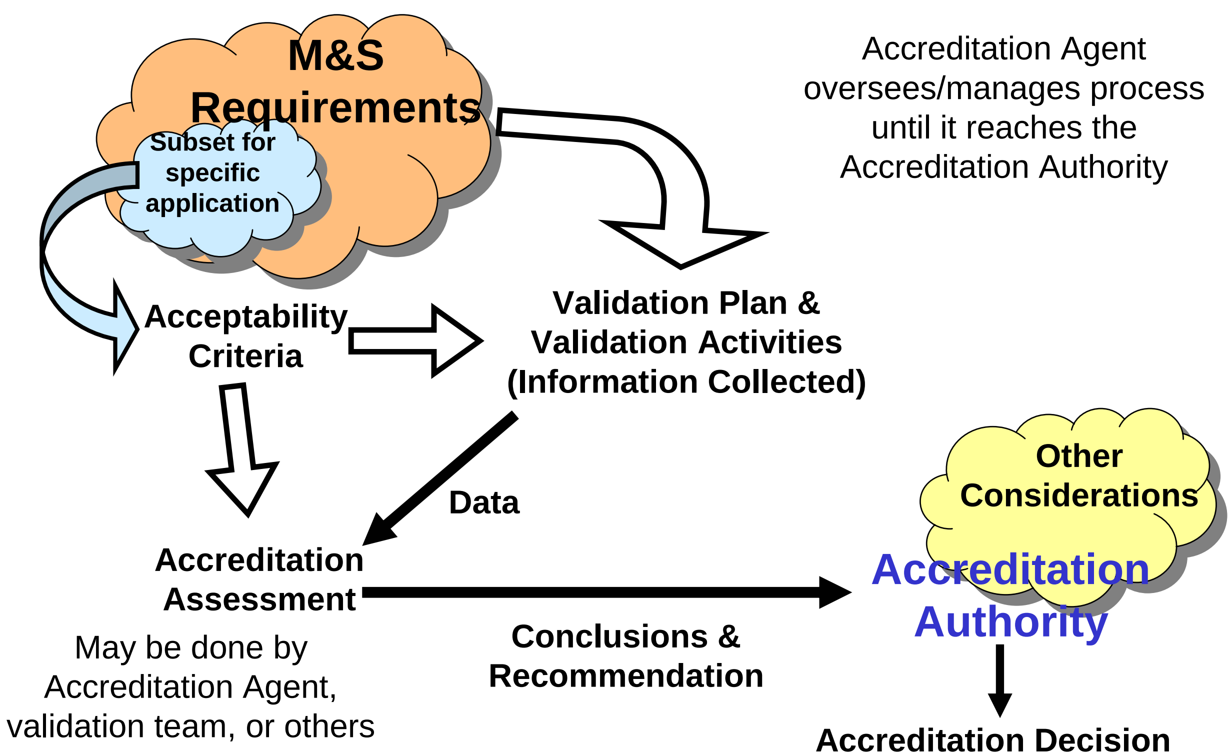


Figure 3.1-1 Accreditation Process in a Nutshell

To illustrate how these factors make referent identification non-trivial, consider the referent of the natural environment for a simulation whose purpose is to predict sensor performance capability. If a radar sensor is involved, natural environmental factors that could be pertinent include terrain (to determine where radar signals would be masked by terrain features), precipitation intensity, extent, and duration (since radar signals can be attenuated by precipitation), temperature and pressure gradients with altitude (these affect radar ducting), etc. These factors all depend upon geographic location, vary with time of day and of the year, and will change from one day/year to the next. For some purposes, a simple free-space radar range computation will be adequate (a computation which is not affected by any of the factors mentioned above); and in other cases, all of these factors must be considered. In some cases, parameter values in standard handbooks or catalogues will provide the information needed. In other cases, detailed observations must be made at the specific site where the radar will be located. And in some real-time applications, the information may have to be current (not historical -- neither recent history from the previous day or ancient history from years before will be adequate).

Intended use of the M&S and the attributes of the reality specified in the M&S requirements determine what characteristics the referent should include (what parameters should be represented and how). Information sources which contain all required characteristics for specified reality attributes provide an initial set of possible referents. Fidelity requirements then determine which information sources are acceptable as a possible referent. For complex M&S, the referent is often drawn from a collection of information sources, not from a single information source.

Some required information may not exist in (or does not have needed fidelity, or fidelity is uncertain) any available referent. It may be necessary to institute a test program to develop the needed information. Another option is to allow a theoretical value or SME estimation to be used for some referent characteristics instead of acquiring higher fidelity test data. As a last resort, M&S requirements may have to be revised so that available information can support validation assessment(s) of a subset of the m/s. In such cases, the validated M&S may address only part of the problem instead of the whole problem. A model-test-model approach may lead to acquisition of additional information that will support validation of additional required M&S capabilities.

Correlating the referent with M&S requirements is essential; otherwise, the M&S may not be able to support its intended use. By analogy, capability to accurately read time from a watch face will not ensure one is on time if the watch does not have the right time. The referent must be able to support M&S intended use.

3.2 Referent Selection

Selection of the referent (where there is more than one possible source of acceptable information) is usually done on the basis of direction, convenience, economics, or a decision to use proxy information for the referent. Any possible referent must be acceptable in terms of scope, reliability, credibility and fidelity.

> The possibility of multiple acceptable referents will be illustrated by considering possible referents for the length of a meter. From 1889 to 1960, the official international referent

for the length of a meter was an international prototype of a meter with an estimated accuracy of about 10^{-6}; from 1960 to 1983, the international referent for the length of a meter was based on the wave length of ^{86}Kr with an estimated accuracy of about 10^{-8}; and since 1983, the international referent for the length of a meter has been based on the speed of light in a vacuum with an estimated accuracy of about 10^{-9}. Often, people use a less accurate surrogate for an official standard referent, anything from the ancient informal measures of the distance from finger tip to nose or a long stride to a measuring tape, yard stick, or more precise mechanical measuring device. Depending upon the application, referents with accuracies of a few percent (or more) could be acceptable. Extremely accurate, and correspondingly expensive, referents based on international standards for length are not required to estimate how much paint will be needed to paint a wall. Measurement accuracies of 10-20% will be more than adequate. This simple example illustrates how a variety of referents may be adequate for a particular application.

When multiple adequate referents are possible for an m/s, it is usually desirable to consider more than one of them before selecting the referent. Once the list is assembled, each of the proposed referents should be analyzed. The analysis should cover the M&S component(s) for which the referent is applicable, M&S requirements and experiment objectives addressed, and information reliability or associated uncertainties. The set of information selected as the referent should address all relevant characteristics and incorporate the least uncertainty relative to M&S results.

3.2.1 Referent Selection by Direction

When the referent selection is based upon direction, the referent selected is often one that the M&S sponsor or user has specified and the accreditation activity/authority has approved. As long as the referent specified in such direction is acceptable, there are no problems with referent specification by direction. However, potential problems can occur when a referent specified by direction is not adequate in terms of accuracy or scope.

Inappropriate referent selection by direction is most often a problem when SMEs have to serve as the referent or supplement facts and theories as part of the referent. Sometimes SME inappropriateness is not apparent. SMEs can be qualified by organizational association as well as by technical expertise [Pace & Sheehan, 2002]. In some cases, technical expertise may be the only consideration in selecting SME as referents.

3.2.2 Referent Selection by Convenience

When the referent is selected for convenience, the referent may be the one easiest to access, one already available vice one coming in the future, or one that the M&S team knows and is comfortable working with. As long as the referent has adequate fidelity, no issues arise from selecting the referent on the basis of convenience.

3.2.3 Referent Selection by Economics

Since cost of information is usually related to its quality (better costs more), it makes economic sense to select the least costly information that satisfies referent content and quality requirements (even if such is not the best quality information in some sense). The key is that the information is adequate in scope and accuracy to serve as a referent. As long as a collection of information satisfies the adequacy test for M&S intended use, it is an appropriate referent. The most appropriate one will typically be the least expensive one. If economic considerations preclude use of an adequate referent, then M&S requirements or intended use may have to be modified before a potential referent can become acceptable.

3.2.4 Referent Selection by Proxy

Very often, it is not possible to identify a referent for the specific performance under study. For example, this will nearly always be the case when the reality represented by the M&S is a future system. In that case, performances of similar tasks on existing systems may be adapted to provide a referent. As an example, when assessing reaction and decision times for operators of a new combat system or machinery control system, there may not be any data regarding operator performance; however, past experience with operator performance in the accomplishment of similar tasks on existing systems may be used as the referent for simulations of the new system.

3.3 A Caveat

Any acceptable referent has to have adequate fidelity if it is to be a meaningful standard of comparison for use with M&S results. It is easy to think of fidelity only in terms of parameter or characteristic accuracy. However, other aspects of fidelity are also of concern for referents: scope, resolution, and context.

- Scope is concerned with the range of parameters or applications of concern for the referent. For example, information about a material, such as H_2O, varies with material phase (solid, liquid, vapor/gas), which is a function of temperature, pressure, salinity, etc. The scope of a referent for H_2O would indicate the range of relevant physical conditions over which it would be applicable.

- Resolution is concerned with the level at which distinctions can be made in the information of the referent. Some information might describe an item's behavior and characteristics at atomic levels; other information at the component level; and other information may be at subsystem, system, or unit levels. In dealing with M&S representing human behavior, the referent might have information about individual behavior, or only about team or group behavior. Resolution of information in the referent must be at the lowest level required to support M&S validation assessment.

- Context addresses the environment within which the referent information is applicable. Sometimes the context is simply a set of assumptions; sometimes the context is a physical condition (such as pressure). There is a difference between context and scope. Context is not specified in the variables used to represent entities, processes and interactions described by the information in the referent, but is concerned with parameters and factors that might influence the measurements (values) of the referent information. For example, the information about how people perform certain tasks that is used as a referent for M&S description of human behavior representation (HBR) probably varies with whether those people were observed by their bosses when their performance was measured, whether the people felt their performance was critical (such as a promotion depended upon it), or other such conditions. Such conditions are part of the context for that information.

4. How to Specify/Describe the Referent

It is important that however the referent information is identified or selected, that it should be thoroughly described, and its role for comparison with M&S results be specified in detail. The referent is the norm against which M&S results are compared for the purposes of validation. Credibility of the referent is directly related to how well the process or phenomenon represented in the M&S is understood.

4.1 Where to Specify the Referent

There are many ways to describe the M&S life cycle. The M&S life cycle can be broken into eight phases, which may be passed through repetitively and several of which may be concurrent. These phases occur regardless of the M&S development paradigm employed, whether a serial like the waterfall paradigm or an iterative one. These eight phases are:

- Requirements (expression of need or desired M&S capabilities)
- Planning
- Conceptual Model ⇐ ***recommended location for referent identification/specification***
- Design
- Implementation
- Test
- Use
- Modification

While the validation assessment referent may be identified as part of the requirements or planning (or even late in the simulation life cycle), it is better to identify it as part of the M&S conceptual model. Referent specification in the M&S conceptual model ensures that the validation assessment referent is available for both conceptual validation and results validation, as well as making the information of the referent available as needed for M&S development. The simulation conceptual model, as specified by the *DoD M&S VV&A Recommended Practices Guide* [*RPG*], consists of simulation context, mission space (representational aspects), and simulation space (implementation aspects). The validation assessment referent should be identified and specified as part of the simulation context. Although the simulation conceptual model is the best place to identify the referent, it is more important that the referent be specified and described as suggested below.

Emphasis upon the conceptual model as the best place to identify and specify the referent for validation assessment may require changes in some M&S development processes. Some M&S are developed without producing an explicit conceptual model. This is unfortunately since lack of an explicit conceptual model has caused many unnecessary M&S development and use problems.

4.2 Referent Specification Dependence on M&S Type and Application

Description of the referent may have to satisfy different needs for applications in various M&S communities. To explore this possibility, the Referent Study identified five varieties of M&S

and five application categories for careful examination in developing guidance for how to describe the referent. These ten kinds of M&S and their applications are not exhaustive (they do not cover all M&S possibilities) and there is overlap among them. However, these M&S varieties will provide enough diversity for insights about how referent descriptions may vary by M&S variety and by application category. An interesting kind of simulation that is not distinctively identified in the ten kinds listed below is what some call "emergence" models ("bottom-up" systems, agent simulations, etc.), such as represented by StarLogo software from MIT's Media Lab or Will Wright's SimCity. We lump this kind of M&S into the category of M&S using adaptive programming. In emergence M&S, there may be behavior rules for how entities in the M&S interact at the micro-level, but the macro-level behavior of the system evolves as interactions occur, sometimes with patterns that were not understood ahead of time (and for which there may not be data available which might be used as a referent). Emergence models have proven useful in biology (such as being able to represent ant colony behavior), traffic, etc. Defense applications of emergence models are also growing. Careful thought about how to identify and describe referents for this kind of model may lead to additional insights about referents, especially since requirements may change with emergence as the simulation is used [David et al].

4.2.1 Reference Dependence on M&S Variety (indicated by a dominant M&S characteristic)

We explore the kinds of information needed for the referent in each of the five kinds of M&S variety indicated in this section. The material in this section should be viewed as preliminary and partial. We expect that future articles and papers will clarify referent descriptive needs for various M&S applications.

4.2.1.1 Extensive use of adaptive programming.

There are many adaptive programming techniques employed in modern M&S: artificial intelligence (AI), knowledge-based systems (KBS), genetic algorithms, fuzzy logic, agent-based simulation, neural nets, etc. Sometimes such an M&S will be labeled as a complex adaptive simulation (CAS). Some of these approaches learn. Some evolve, and change the M&S structure and algorithms. Some M&S with adaptive programming are designed to demonstrate how things behave where their behavior is unknown. There are no standard information sources for referents for such M&S at present. Sometimes referents for low level, detailed component/unit behavior or characteristics can be identified and defined; then M&S results are used to determine aggregate level behavior. Some will use aggregate level information to set a pattern that is desired and adjust M&S parameters until that kind of aggregate performance is obtained.

However the referent information is selected, it should be specifically identified and thoroughly described, and its role for comparison with M&S results be specified in detail.

4.2.1.2 Extensive use of human behavior representation (HBR).

M&S with extensive HBR are intended to create realistic or credible human responses by computer generated actors. Perhaps the ultimate HBR referent is the Turing Test [Turing, 1950], but how to specify such is still elusive. "As a discipline in simulation technology, HBR validation is still relatively immature with no theory, few tools and techniques and considerable but poorly documented experience. Two sources of information establish a firm foundation for the advancement of HBR validation technology, a broad experiential base and a more mature related field, knowledge-based system (KBS) verification and validation. HBR validators have learned many lessons from existing and future systems that deal with requirements, the subject matter expert-software engineer process, association of the HBR with the synthetic natural environment, and documentation. These lessons supply a rich source of guidance for future HBR validation activities. KBS verification and validation is considerably more mature with a very large literature base to support it. HBRs, as a form of KBS because of their knowledge bases, can benefit significantly from this technology base. With these resources and increasing realization of its importance across the broader simulation world, HBR validation is poised to mature very rapidly." [Harmon et al, 2002]

Not all AI or KBS M&S are adaptive. Some are just rule based. Sometimes such processes are used to mimic human behavior in M&S. When that is the case, the validation referent is human performance. Of course, it is best if the human performance is appropriately described for its variations and dependence upon conditions. However, this kind of referent has a particular characteristic. It restricts assessment of representational goodness to the best that a human can do. If the objective of the M&S is to represent human performance, then that is exactly what one wants as a referent. However, if one is trying to determine who well a function (such as pattern recognition) is performed, the referent may be as limited as trying to determine the precision of a measuring device by comparing its capability with that of an ordinary ruler. At least in theory, an automated process may be able to perform some functions much better than humans, and use of human performance as a referent is thereby very limited in its utility as the validation assessment standard.

The following section on Distributed Simulation, Section 4.2.1.3, has extended comments on humans and HBR in distributed simulations that are germane to HBR in general.

4.2.1.3 Distributed simulation

Distributed simulations pose a special validation assessment problem in that they are intrinsically more complex than many unitary simulations because they involve both the interaction of multiple processes and the infrastructure which makes the interaction possible. A coarse assessment of the simulation's performance may be obtained if the results can be compared to referent information consisting of a previous "real life" event, either from an operational event or in a test/training situation. Because of limited controls in such real life events, there is significant variability in data from "live" events. This places limitations on the ability of real life events to serve as a referent as a result of the real life event's intrinsic variability.

Problem decomposition is one of the basic procedures employed in scientific methods. It can also be applied to the referent for a distributed simulation, as described below. It is a particular application of what some call "piecemeal validation." Simpson's Paradox (the potential problem of the direction of a conclusion from examining pieces being at odds with the direction of conclusion for the whole, see Pearl, 1999) is always a concern for piecemeal validation and should also be a concern in the decomposition approach to the referent for the distributed simulation.

A more viable process (than real life events as the referent) is to decompose the simulation federation into its functional components. Each component may then be validated individually. This permits a sensitivity analysis and comparison against theory and historical data. At the function-level, other models may be available that have comparable or greater precision and that have a prior validation history. Also at the function-level, more options are available for the selection of a referent. Once the individual functions have been validated, their interactions among the simulation's elements are assessed. This is accomplished by the examination of the interfaces between each of the components. The V&V agent must follow the logical threads across the interfaces to ensure that the data being received is, in fact, what was expected. While many algorithms use the same terms and produce results with the same units, they may have underlying assumptions or approximations that unacceptably skew the results when they are utilized in subsequent processing.

At the component level, referent information is often more readily available. It may take the form of a high resolution representation of the phenomenon that is being treated within the simulation by the component. Or a general-purpose simulation engine may be available that can be used to prepare a "reference" representation of the component being validated. A third source of referent information could be a data set from an experiment in which the phenomenon was observed under laboratory (or controlled) conditions. A fourth source can be data from field observations or a live exercise; because it has limited control, the errors associated with this type of data are somewhat greater than those observed from controlled situations. A final source of referent information is the subject matter expert (SME); the SME is often assigned a larger error range than other methods; however, for complex simulations, the SME can often account for more of the uncertainties using experience than can be thoroughly bounded by other referent information.

Although components are executed in separate but linked processors in a distributed simulation, the use of referent information follows the same process as for a simulation that is executed on a single processor. The potential pitfalls come from issues related to timing, execution coordination and processing delays. In addition to the referent information associated with the individual algorithms, the referent must be selected that demonstrates that the data flow within the federation is not adversely affected by the timing associated with data transport among, or between, the federates.

Referents for Humans, Software, and Systems in Distributed Simulations. We discuss referents for humans, software, and systems in the context of distributed simulations. Additional discussion of human behavior representation (HBR) may be found in Section 4.2.1.2.

Human Factors. Probably the greatest uncertainty in simulation performance is encountered in the representation of human performance. The following should be noted:

- The "average person" is not the sum of the "average" body parts. There are significant variations between the genders; in addition to size, one finds differences in reach and mobility. These have great significance when modeling the physical properties of the human body; and when simulating its motions.

- Differences in decision and reaction times have been observed between members of the Army, Navy and Air Force performing the same task in operational situations in which all had completed the same training for the task and had comparable experience. This has been ascribed to differences in the underlying command structures and operational doctrines of the services. This is a specific example of “context” (as discussed in Section 3.2.5).

- Age, experience, outside distractions and fatigue can cause significant variations in human performance. These factors must be taken into account in the simulation of Command and Control (C2) systems and in the modeling of equipment operation.

When selecting a referent for human performance, the first criterion is that the data sets should be as representative of the simulated population as possible. The second selection criterion is that the task represented by the data set should be representative of the action to be taken—motion or decision. Note that there can be significant differences in timing between training and exercise conditions and actual operating conditions; however, as proficiency increases, the two values will converge.

In situations in which data sets are not available, texts and industry consensus standards can be consulted for “typical” or “standard” values for similar tasks. These data may be use to calculate a referent value for the simulation. Care should be taken to ensure that the sources selected for the referent are independent from those selected by the simulation developer.

Software. Software can be used as a referent. It may take two forms. First, a software package that has previously been accredited as representative of the system being simulated may be used as a referent or comparison standard for the simulation being validated. Second, a specialty or general purpose software package may be used to construct a reference model of the system being simulated.

The use of accredited software packages for the validation is particularly applicable when complex systems are being simulated and detailed models have been developed and tested. An example of such a situation is the incorporation of a representation of the Navy’s AEGIS Combat System or the Army’s PATRIOT Missile System into a larger simulation federation. In such a case, a model provided by the program office could be used as the referent for the simulation being validated.

For notional or developmental systems, reference models can be developed using commercial software packages. Examples of such packages are COMNET® for communications systems and Extend® for process modeling. Care should be taken when constructing reference models for use as referents. First, the precision and accuracy of the models should be sufficient to be a satisfactory predictor of performance adequacy. Second, the model can only reflect the performances specified in the requirements and the characteristics identified to the point that the model was constructed. In this situation, the model, used as a referent, really represents a second solution to the system being simulated. Accordingly, many uncertainties may remain; and the V&V Agent should attempt to identify them.

Real Systems. Often, an operational system is being abstracted into a simulation federation. When this occurs, data from representative tests and evolutions may be used to validate the simulation. Care must be taken to ensure that the simulation set-up is the same as the one used for the operational system at the time the data were taken. However, the availability of data from the operational system can be the most authoritative source.

The next level of abstraction comes when an operational system's capability is being extended. In that case, the simulation's performance is an extrapolation of the performance data for the operational system. In that case, extreme care must be taken to ensure that the new or modified functionalities are correctly modeled in the simulation. In this case the procedures proposed for reference models developed from commercial software packages should be followed.

4.2.1.4 Extensive use of aggregation in the M&S.

The standard referent for M&S aggregation is based upon more detailed representations of the entities, processes, and phenomena represented in the aggregated M&S. These more detailed representations typically come from "higher fidelity" M&S. "Higher fidelity" in this context means that the actors, systems, entities, interactions, processes, or environments in the M&S are represented in more detail and at a finer granularity of resolution (for example, with individuals instead of groups as the smallest level of distinction). As noted by Bigelow and Davis [2003], such "higher fidelity" M&S do not always give better results in terms of aggregate level behaviors. Specification of the referent context is particularly important since there are often differences between the context (assumptions, conditions, etc.) of the aggregated M&S and the context of the higher fidelity M&S that can impact the appropriateness of the referent for the intended application. Use of results from "higher fidelity" M&S as referent information for aggregated M&S validation assessment brings the potential problems of Simpson's Paradox [Pearl, 1999] into play. A quality referent should indicate how this problem should be addressed.

4.2.1.5 System/hardware/software/human in the loop M&S

Many of the referent issues for this kind of M&S were covered in Section 4.2.1.3 on Distributed Simulations.

M&S can be used for understanding, describing or predicting the behavior of a single item or for items from a population of items. A flight simulator, for example, can be used to predict the likely performance of a particular pilot (or crew) in flying a particular aircraft. This is an example of a "single" item in the first sentence of this paragraph. If the same flight simulator is used to estimate performance of pilots (or crews) for a particular kind of aircraft, then it is an example of "items from a population" in the first sentence. In that case, the personnel involved may not be a good referent. The personnel might be aces or duds, or otherwise not representative of behavior and performance variations of the population of interest. Thorough

understanding of the population of interest, its characteristics and variations, is essential in determining what should be the referent for this kind of M&S. Typically that issue is ignored, and often a single item is blithely assumed to be representative of the population of the item.
characteristics and variations, is essential in determining what should be the referent for this kind of M&S.

Referent specification for this kind of M&S should indicate whether intended use is for a single item or for items representative of a population. If it is a single item, then the relationship of that item to the referent should be clear. When the referent is the item itself, then variations in the context for the item in the M&S should be clearly delineated and their impact estimated (such as the "white knuckle" reality of real flight versus experience in a simulator, or the lack of vibration and acceleration forces in a missile seeker in a HWIL simulation vice one in actual missile flight). If items representative of a population are intended, then the referent should specify the characteristics and variations of the population and the relation of the items used in the referent to that population.

4.2.2 Reference Dependence on M&S Application Category (indicated by general application domain characteristics).

4.2.2.1 Computational science and engineering applications (such as CFD M&S).

The computational science and engineering M&S community is focused on experimental data (from particular tests) or in some situations from standard benchmark cases as referents for validation assessments. In that community, "benchmarks" are usually part of verification activities. "Benchmark solutions refer either to analytical solutions, *i.e.*, exact solutions to the PDEs with the specified initial conditions and boundary conditions, or to highly accurate numerical solutions. However, we believe that in the solution of nonlinear PDEs or solutions with discontinuities or singularities the most reliable benchmark solutions are analytical solutions. In validation activities, accuracy is measured in relation to experimental data, *i.e.*, our best indication of reality." [Oberkampf et al, 2002] Some in this community reject use of the term validation for comparisons with results from other simulations (no matter how accurate those simulations results are perceived to be) or for comparison with theoretical predictions.

4.2.2.2 Engineering level applications.

Engineering level M&S tend to have resolution at the component or sub-component level and use algorithms that describe item characteristics and behavior at that level or higher. Typically the information used as the referent in validation assessments are drawn from specifications and requirements (for future systems or items), from tests and other data (for existing systems and items), or from higher fidelity M&S (such as computational science and engineering m/s). Key aspects of referent specification are: identification of information sources, description of how information from different sources will be combined, quantification of variations and uncertainties in the information, and delineation of the information context.

4.2.2.3 Game/training applications.

The basic referent for an M&S training application is human behavior, performance and skills or capabilities obtained in the reality represented by the M&S. For example, in driver training, behind the wheel (BTW) situation becomes the basic referent, as described in the Federal Motor Carrier Safety (MCS) Administrations validation of simulation technology for training, testing, and licensing of tractor-trailer drivers [MCS, 2000]. Unfortunately, the situation is not that simple since it may be that M&S-based training (whether combined with BTW training or just by itself) could train drivers better than all BTW training, as illustrated in simulator experience in the Netherlands [TNO Human Factors, 2004]. In this case, there may not be a well defined referent. Standard performance tests, as demonstration of driver skill, can be specified as a referent for the skill acquired, and longitudinal tracking of driver crashes, traffic citations, supervisory ratings, etc. may be used to compare driver safety as a factor of the kind of training received, but none of these approaches goes far toward precisely defining a referent to be used in M&S validation assessment. In many cases, SME judgment is the referent, as in efforts to validate medical training simulators reported at the 2002 Army Science Conference [ASC, 2002]. Typically the referent when SMEs are used in this way comes from a number of tests with the training simulation.

The referent for validation assessment of games is even less clear than for training M&S, as illustrated by simply reviewing articles in any of the issues of the *Simulation and Gaming Yearbook* since its beginning in 1993. If one considers games with an educational aspect as a particular kind of training M&S, the referent issues above apply. If one focuses on games for entertainment, then the referent is basically SME judgment about what will appeal to players, with post-development experiences of players and their satisfaction/interest in the game indicated by sales, play, or surveys used as confirmation (or abrogation) of the SME assessment.

4.2.2.4 Military theater-level/campaign M&S.

The validation of theater-level and campaign models and simulations is one of the most vexing problems facing V&V personnel. This is due to two characteristics. The first is the inherent complexity of the system being modeled. The second is the number of uncertainties that must be dealt with.

Complexity is often handled by a process of “partial differentiation;” specifically one factor is either held constant or allowed to have 100% availability. Examples of this include the assignment of 100% logistics availability, “infinite” ammunition, or ideal weather. Once all of the other variables have been assessed, then the variables that were artificially constrained are allowed to vary over permissible ranges. This process permits the analyst to define the solution space for the model or simulation and ascertain the sensitivity of the outcome to the performance of each of the variables.

With respect to the number of uncertainties, validation must rely upon statistical methods of sampling and the ability of the developer to quantify the statistical distributions and any of their associated dependencies.

The foregoing poses many problems in the identification and selection of referents. Four basic methods are employed, usually in combination.

- One method of resolving the problem is to compare the results of the larger (theater-level) simulation against more detailed but restricted (engagement-level) simulations. However, great care must be taken to ensure that the simulations employ the same context. Often the theater-level simulation and the engagement-level simulations are focused on different issues (system measures vice unit measures) which can preclude valid comparisons of their results. When such differences do not preclude appropriate comparisons, the results from the more abstract theater-level simulations may be shown to be consistent with detailed simulations that in some cases can be traced back to more elementary engineering considerations, what some call "the physics" of the problem, or "first principles."

- A second method is to disaggregate the larger simulation into operational or functional areas—logistics, weapon effects, maneuver and so forth. Then the results of detailed simulations of each area can be compared with the performances found in the simulation being validated.

- The third method examines the simulation's performance in terms of the comparability of its results to other simulations of greater or lesser detail and to live test data.

- The fourth method is reliance upon SME judgment that the simulation's performance complies with SME expectation.

The following example illustrates some of the ways these different approaches interact. The performance of a missile defense system can be assessed using the Extended Air Defense Simulation (EADSIM), the Extended Air Defense Testbed (EADTB) or the Theater Missile Defense Systems Exerciser (TMDSE). EADSIM is a high-level simulation that uses databases to characterize system performances. EADTB is a detailed simulation that explicitly simulates the operation of every tactical data processor and entity on the battlefield. TMDSE is a hardware-in-the-loop simulation that links operational systems in common operating environment. In operation TMDSE is able to link one or a few units of any one type in a federation. Accordingly, it can examine the interactions between the various tactical data processor types; however, it cannot assess the interaction of large numbers of similar tactical data processors operating together (i.e., AEGIS ships, or PATRIOT batteries). EADTB is not limited in the numbers of units that it can represent on its "game board;" however, because it explicitly simulates each tactical data processor, it is constrained by the processing time required for each run. That establishes a practical limit upon the number of units or the complexity of the problem. Finally, EADSIM is able to simulate very large numbers of participating units because of the simplification of its algorithms.

In terms of the referent, the goal is to show comparability among the results from each of the simulations when running the same scenario. When going from the results obtained with TMDSE, an extrapolation is being made from "one to many" with operational systems in a test environment. In the case of EADTB, the extrapolation is "many to many" with great detail in a simulated environment. For EADSIM, the interpolation is "many to many" with low resolution or limited detail.

In the world of operational testing, the availability of operational units and range resources is limited. Accordingly, a test of a single unit must be extrapolated to assess its performance in an operational scenario. Using the tools described above, the process would take the individual test and compare it to the results from TMDSE to assess its performance with other systems. Those results would then be compared to the results from EADTB for detailed performance in a limited operational scenario. Finally, the results would be compared to the performance in EADSIM for a high level assessment in a large-scale operation. In this case, we have a series of referents. Each traces its validity back to a prior event. The keys to success for this method is configuration management of the individual simulations and test platforms, verification and validation that can be traced back to operational and test requirements, and sufficient test data to permit identification an sensitivity analysis of experimental uncertainties.

It is important to specify how the possible sources of information that will be used for validation assessment of the simulation are going to be combined and used.

4.2.2.5 Non-physical applications (such as economic M&S).

The primary source of referent information for non-physical M&S applications is historical experience or evidence related to the reality represented in the m/s. Regression techniques are frequently employed to show how M&S results correlate with the historical information.

It is important that such historical information be decomposed in all areas of interest for the M&S application, and that correlations be performed in all areas to ensure that the M&S can support its intended application. In the mid-1960s, a very large linear program was developed as a tool to assess impact of various factors on Gross National Product (GNP). It had thousands of variables (possibly it was the largest linear program in existence at the time), and was intended to support assessment of many kinds of government polices. Interest in it faded, however, when it became quite clear that when fed historical data its overall prediction of GNP for a few years into the future from the time of that historical data was even less accurate than a simple linear extrapolation of GNP change.

4.3 Basic Principles for Referent Description

Referents have five fundamental attributes: context, domain coverage, attribute distinctiveness, parameter uncertainty quantification, and information coherence. Each of these is discussed below. Some general characteristics should apply to all referent specifications. **Identification** and **specification of a referent** should be **definite** and **unambiguous**. Just doing this consistently will bring improvement over current M&S practice. If some information in the referent is to use only part of the information in a particular source (such as results from a test or set of tests), what is not to be used should be clearly and specifically delineated. For example, if the flight profile of a missile is to come from a particular test, but the radar signatures of the missile in that test are not appropriate for the missile referent, that should be noted explicitly and clearly in the referent identification and description. This is to eliminate any possible confusion about what is to be used as the referent.

4.3.1 Context

It is always important for a referent to specify its context. The context will range from information about conditions under which human performance information (as discussed earlier in Section 3.2.4) is collected to physical conditions (such as temperature, pressure, radiation, etc.) that might influence measured parameters in the referent which are not addressed specifically by information in the referent. Care must be taken to avoid specifying the referent too narrowly. For instance, if an M&S relies on physical characteristics of water, the referent should be as broad as possible so that it catches erroneous situations such as water being in the wrong phase during an explosion or other event, which while rare, might be exactly the sort of unusual condition for which the model's prediction is most useful.

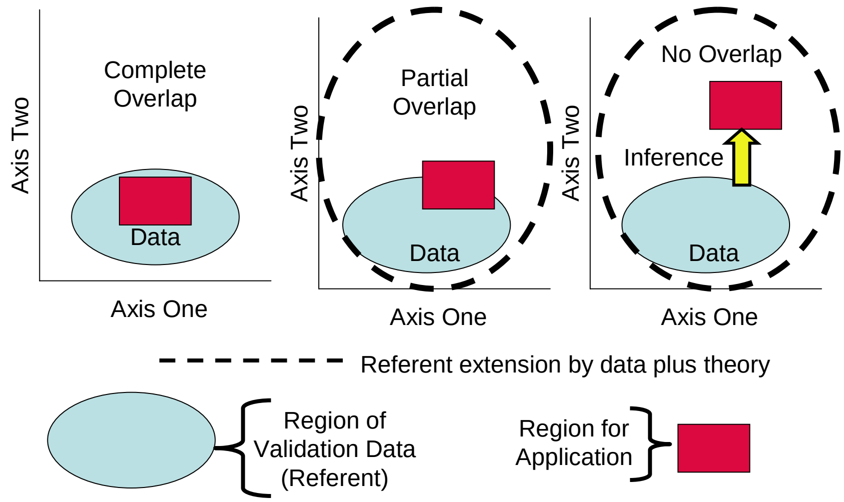
4.3.2 Domain Coverage

A simulation has a specified application domain. Its intended use determines the appropriate application domain. The application domain is always multi-dimensional. Referent description should indicate what portion of that application domain is addressed by the referent. For example, test data used as a referent may reflect steady state, smooth, undisturbed flow conditions for a parameter (such as fluid volume passing through a pipe, traffic on a road, time delay waiting for a technician in on-line support, etc.) but is not appropriate for use as a referent in transition, turbulence, disturbed flow conditions (such as might be experienced if an obstacle were in the pipe or on the road, or the on-line support shift is short handed).

Limitations in referent domain coverage forces one to consider inference. Figure 4.3.2-1 illustrates three possible relationships between data for a referent (the circles) and intended use (boxes) in the application domain. If the referent is restricted to data, only when the data completely overlap the application area can there be high confidence in a quantitative assessment of the relationship between simulation results and the referent. In the other two situations (partial overlap and no overlap), indeterminate uncertainty is present for applications outside the data region. If reliable theory exists, then either of these cases might be reduced to the equivalent of the complete overlap case.

Figure 4.3.2-1

Possible Relationships of Validation Referent Data and Application Domain



General parameter regions are fairly easy to understand. For example, if a simulation application is to address what happens when bodies collide (as with a missile defense kill vehicle hitting its target), the range of interest for collusion velocities may vary from nearly zero (as could happen in a scenario in which the kill vehicle approaches from behind the target) to very fast (10 kM&S or more for a fast interceptor against an ICBM in a head-on encounter). Data from full-scale tests may be very limited, not only in the number of tests, but also in the portion of the speed regime for which there are tests and in the availability of precise information from the tests. Data from surrogate tests (such as sled-tests or light gas gun experiments) may supplement the full-scale test data, but they introduce uncertainties into the data because of test artifacts (such as need to scale results, differences between the surrogate and the real object, etc.) and such tests may not fully cover the parameter regime of interest. Then theory or perhaps very high fidelity simulations may be used for "data" for parts of the parameter regime that testing (either full-scale or surrogate) cannot address. And finally, expert opinion may be used to fill in any remaining information gaps in the domain (and to reconcile any discrepancies among the various kinds of information mentioned).

Description of the referent will indicate how information about the M&S actors/systems/entities, processes, and environments throughout domains of interest are addressed. What information sources are to be used for which portion of the domain, where gaps may exist, and how information from different sources may be reconciled and are to be combined.

4.3.3 Attribute Distinctiveness

The referent is concerned with actors/systems/entities, their interactions, processes, and environment(s) of the reality represented in the simulation. As illustrated earlier, not every possible attribute of these is significant for the referent. All attributes needed to satisfy M&S

requirements should be specified for the referent, and described fully for the domain coverage required with all pertinent variations indicated. For example, if the object size or color changes with temperature, etc. and the object's size or color is important for intended M&S use, that attribute trait should be specifically noted.

4.3.4 Parameter Uncertainty Quantification

There are two uncertainty dimensions in fidelity and validation assessments. One dimension is M&S uncertainty. These arise from imperfect algorithms, computation characteristics (such as table look-up errors), input errors, etc. Often when simulation results are compared with a standard (theoretical curve, test data, etc.), it is assumed that all error or uncertainty is a result of M&S uncertainties. This usually is not the case. The other uncertainty dimension in fidelity and validation assessment is uncertainty in the referent. Only when referent uncertainties are fully characterized can M&S uncertainties be fully characterized[3]. When both uncertainty dimensions are fully characterized, then fidelity and validation assessments can be performed rigorously and fully characterized.

In the computational science and engineering arena, a great deal of attention has been given to quantifying uncertainties, both M&S uncertainties and referent uncertainties, in papers [e.g., Oberkampf et al, 2000], short courses [e.g., Ghanem & Wojtkiewicz, Uncertainty Quantification at SIAM 2003 Conference on Computational Science and Engineering], conferences [e.g., SAMO, 2004], and workshops [e.g., CIMMS, 2002]. Various validation metrics have been suggested in this arena, guidance in regard to validation experiments has been developed so that referent information with characterized uncertain may be known, etc.

A proper referent description will include specification of parameter uncertainties in referent information.

4.3.5 Information Coherence

As indicated in previous discussion, information for the referent may come from multiple sources. Some will be redundant (same parameters for the same part of the same domain), some will be supplementary (different parameters for the same part of the same domain, or same parameters for different parts of the same domain), and disjoint (different parameters and different domains). Some information will have parameter uncertainties quantified, and other information will not. Information coherence is concerned with how information is combined so that information about a particular aspect of the referent (parameter, actor, system, entity, process, environment, etc.) makes sense. Reference description should explain how information combination is done so that information coherence is achieved.

[3] Some feel "fully characterized" is redundant, that it means no more than "characterized." Others worry that addition of the term "fully" provides an excuse for people not to try since it is too hard or impossible to "fully" characterize the uncertainties in an absolute sense. What is intended by "fully characterized" is that adequate information is provided so that meaningful quantitative statistical statements can be made about the relationship between the referent and M&S results with high confidence.

The two examples below illustrate some of the ways this can be done. Neither example tries to show a best or preferred way for combining information.

> Case 1. Information about parameter x will be taken from three tests, all of which address how the parameter varies with respect to y over the same set of values for y. In Test 1, uncertainties in the measurement of x are fully characterized. Tests 2 & 3, which were done at the same test facilities using the same equipment and test personnel, do not characterize x uncertainties nor describe environmental conditions fully. Options for referent information about x include:
>
> - Use all data from Tests 1-3 and ignore uncertainties about x
> - Use only data from Test 1 and fully characterize x uncertainties
> - Use a plot of x from Test 1 with uncertainties indicated, overlaid with data from Tests 2 & 3, as a basis for an equation to characterize x vs y
>
> Case 2. Information about parameter x is available from Test 1 for a portion of the domain of interest. Information about a parameter y, which it is believed that x is proportional to, is available for a portion of the domain of interest not covered by Test 1 (but without any information for the part of the domain that is addressed in Test 1). Two SMEs gave widely divergent opinions about what multiplication factor should be assumed in the proportional relationship between x and y. Options for referent information about x include:
>
> - Use only data from Test 1
> - Use both data from Test 1 and a band for x determined by proportionality to y (with the two SME estimations setting the edges of the band)
> - Declare that x is unknown

A proper description of the referent will explain how information from different sources is to be combined coherently. Obviously general guidance for how information should be combined is beyond the scope of this report; but it is clear that explicit description of how such information will be combined is an important part of a referent's description.

4.3.6 M&S Variety/Application Dependent Aspects of Referent Description

All referent identification and specification should be definite and unambiguous, with clear indications of information sources and how information will be combined with information is drawn from more than one source. The following two sections separate referent description characteristics that depend upon M&S variety or application from referent description characteristics that do not depend upon M&S variety or application domain. The five element paradigm from above for referent description (context, domain coverage, attribute distinctiveness, uncertainty quantification, and information coherence) will be used to organize discussion in the two sections below.

4.3.6.1 Synopsis of Referent Description Characteristics Applicable in All Cases

Pertinent physical aspects of the referent context apply in all cases.

Specification of the extent of the application domain over which referent information applies is applicable in all cases.

Attribute distinctiveness descriptions are applicable in all cases.

Methods to be used to obtain information coherence are applicable in all cases.

4.3.6.2 Referent Description Characteristics Dependent Upon M&S Variety or Application

Pertinent psychological, strategic, doctrinal, tactical, and procedural aspects of the referent context only apply to M&S which involve people, HBR, or automated forms of decision making.

Parameter uncertainty quantification is mainly restricted to parameters that are measured physically. In general, aspects of the referent that are based upon human judgment or opinion are not amenable to uncertainty quantification.

How the referent evolves with M&S use is only applicable with adaptive m/s.

5. Referent Use in Validation and Accreditation Assessments

In this section, the use of referents in M&S validation and accreditation assessments is discussed. The importance of proper identification and description of the referent is emphasized. Three basic situations are discussed. In the first, referents are data-rich (the information about the simuland is abundant as well as necessary and sufficient from fidelity considerations). In the second, referents are data-poor (there is little information about the referent). And in the third, the referent is poorly defined, poorly identified, or otherwise inadequate.

This section also discusses the important relationships between the validation and assessment referent and information used for M&S development and information used as inputs in M&S runs.

5.1 Data-rich Referents

A data-rich referent is always the most desirable situation, especially if uncertainties about parameters of the referent are well specified. Such referents provide the basis for the most reliable and most credible validation assessments of M&S results since they provide an objective and factual basis for statistical comparison of results. However, even in situations with data-rich referents, care must be taken to ensure that the information used for the referent is truly pertinent for M&S intended use. For example, if human size is part of the referent (as might be pertinent for M&S concerned with passenger movement in a new vehicle design), one must be sure that the information about human size is recent since (at least in America) humans are bigger now than they were a couple of generations ago.

In a data-rich environment, it should be possible to separate the information used for the referent from information used to designed and develop the simulation so that M&S validity and capability for reliable predictions can be assessed more robustly. The need for statistical independence between information used for M&S development and information used for validation assessment was mentioned earlier in the report and is discussed in Section 5.4 below.

5.2 Data-poor Referents

In a data-poor situation, the referent is mainly theoretical by necessity. The theory may be explicitly articulated and well-formulated, as in an astrophysical simulation of processes inside stars [e.g., Calder et al] for which we have only inferred data, no *in situ* observations. On the other hand, the theory may not be well-formulated, as is often the case when SMEs are used as part of the referent. The SME may not be able to articulate how a judgment was reached or upon what the assessment is based. The credibility of validation assessments based upon data-poor referents is always severely limited, but a thoughtful probing of SMEs (and good documentation of their insights) makes the best of this situation.

In data-poor environments, it is seldom possible to separate information used for M&S development from information used for validation assessment because all information has to be used in M&S development. Some, especially M&S personnel within the computation science and engineering community, would claim this limits M&S assessment to calibration and prohibits validation of M&S predictive capabilities. It is recommended that the referent description/specification note that referent information was also used for M&S development so that appropriate caveats may be associated with the validation assessments. This report does not attempt to prescribe how such caveats are determined, only that the lack of statistical independence[4] between information used for M&S development and assessment be noted so that caveats can be indicated with the assessment as appropriate.

5.3 Inadequate Referents

In the third case, the referent is not clearly identified or has recognized inadequacies (such as contradictory information, so that if simulation results agree with some of the “referent” data, they will not agree with other parts of the “referent” data). If there is no clear way to sort the referent information so that it is coherent and non-contradictory, then it is necessary to declare that validity assessments are not possible for the simulation because of referent inadequacies. However, as Hodges and Dewar noted a decade ago, even a model which cannot be validated can have utility and value [Hodges and Dewar, 1992]. In the data-poor situation discussed in Section 5.2, the use of theory or SMEs as the referent allows one to make fidelity and validation assessments, even though the credibility of such may be low; but when the referent is contradictory, one cannot even make a low credibility assessment. Inadequate referents challenge the courage and professionalism of V&V personnel since they force the V&V personnel (if competent and honest) to declare to M&S sponsors, users, and others that validation assessment is not possible because of referent inadequacies. Most V&V personnel are acutely aware of how often the messenger bearing bad news has been shot; many have scars that demonstrate the hazard of bearing bad news.

5.4 Referent Relation to M&S Development and Run Information

Information about the reality represented in the M&S is used to develop the M&S, in running the M&S, and as the referent in validation and accreditation assessments. It is reasonable to inquire about the relationships among these three sets of information since they all deal with the same reality.

In a data-rich environment, it is most desirable that the set of information used for M&S development, the set of information used as inputs for running the m/s, and the set of information used as validation referent be statistically independent. This is particularly important when M&S results are used to predict how things will be in regions for which data are sparse or absent. The rationale for this is simple. This approach provides the highest likelihood that M&S results will be most representative of the reality represented in the m/s.

The medical community has found that conclusions from clinical trials can be vary significantly from what is believed to be more correct if appropriate care is not taken in control of such

[4] By statistical independence, we mean that the information is drawn from separate sets of data; the term usually implies physical independence too. For example, measurement of the same object using the same instrument by the same person on two different occasions would produce two data sets without physical independence. Measurements of different examples of the same kind of object or by different instruments or people would have physical independence as well as different data sets.

statistical issues [Kunz and Oxman 1998]. Similar concern is needed in M&S assessments. Many do not appreciate the issues associated with using the same information for M&S validation assessment as used for M&S development.

Concern about this general problem, and its specific related problem of not letting a modeler know experimental outcomes before the model describing the same situation is run, is abundant in the computational science and engineering community. The need for the modeler to know the exact conditions of the experiment precisely before running the model is understood. Numerous guidelines are presented for the ways that "empirical adjustable parameters" [Roache, 1998], *knobs*, *dials*, *fudge factors* in more colloquial terminology, must be treated for M&S results to be acceptable in peer-reviewed circumstances.

In data-poor environments, statistical independence among the three sets of data may be impossible. There simply may not be enough data. The three data sets may even have to be identical. This condition limits what can be claimed about the validity of simulation results. One can describe with quantitative precision how well the M&S reproduces its input data, but one cannot make very meaningful comments about M&S predictive capabilities. In that regard, one may be unable to make a stronger assessment than one can with an inadequate referent. Candor about limitations of validation assessments in such circumstances is an important aspect of V&V professionalism. The words of statistician George Box [1979], who at the time was the Past President of the American Statistical Association and President of the Institute of Mathematical Statistics should not be forgotten in this respect: "all models are wrong, but some are useful."

6. Referent Research Needs

There are three primary research needs relative to referents. The first is concerned with how to consolidate specified referent information so that it is coherent and so that its uncertainties are appropriately specified. Uncertainty quantification efforts within the computational science and engineering community [e.g., Ghanem and Wojtkiewicz, 2003] are basically addressing this problem within their domain, and many of the mathematical techniques may have some applicability elsewhere too. It is unlikely that the approaches identified for the computational science and engineering community will deal with all issues of concern in other M&S communities. For instance, some communities rely more on proxy information, and there is a need to develop methodologies for assessing the adequacy of proxies in referents. There is also a need to develop concepts for evaluating the overall adequacy of a referent, where "adequacy" can be a very slippery and subjective notion.

For example, in hardware-in-the-loop simulations, what is the appropriate referent for the missile seeker (or other item of hardware) in the loop? Should the item itself be considered the referent? If so, are there any variations associated with other examples of that hardware item that need to become part of the referent description (some hardware used in such simulations was hand-built in a laboratory and did not come off the normal production line, for example). Ways to determine and describe the representativeness of such hardware is part of this research topic. Likewise, how to determine referent information for people who are involved in human-in-the-loop simulations and war games (or other games) is also part of this research topic [some of the issues were mentioned in Section 4.2.1.3].

More generally, there is a need for codifying best practices for the use of SMEs and other sources of qualitative information so as to achieve the maximum rigor in referents that are based on such information. Good practices from the social sciences might be adapted to regularize the use of such inputs. It is also necessary to develop methodologies for good sensitivity testing of referent data so as to understand which aspects require the strongest data in order to properly illuminate M&S quality, and which have lesser affects on the VV&A. Part of this effort will require research into appropriate ways to combine descriptions of variations (uncertainties) in referent characteristics for the different actors/systems/entities, processes, and environments that comprise the referent.

The second primary research need is concerned with dynamic referents. If a simulation is to represent something that changes with time or other parameters, how can a referent at a previous state/condition (which may be the only situation for which data are available) be used with simulation results for a subsequent situation, especially if emergent behavior or adaptive processing is used in the simulation? This issue will be an increasingly important problem for simulations in the future. For example, behavior in many biological systems and in crowd/traffic situations varies with population density in ways that are not easily predictable by algorithms at the macro-level, though behaviors can be reasonably described at micro-levels. Should the referent be restricted to micro-level behavior? Or can there also be some kind of useful referent at macro-levels?

A third area of research addresses the method for documenting the use of referents in M&S validation and accreditation. Specifically, how is the referent documented so that it is traceable and reproducible? The key concern is M&S reuse or the incorporation of M&S results into subsequent M&S. How can we achieve a better understanding of how to use results of one M&S to inform another—e.g., when and how to integrate fine-scale models into coarse-grained ones? In a particular case, if we want to use an accepted simulation as a referent for another M&S, we need to know how differences in scale, fidelity, and assumptions of the "referent simulation" should be interpreted as part of a VV&A process for the M&S being validated.

7. Conclusion

This report indicates the importance and complexity of validation assessment referents. Its treatment of this subject should stimulate fuller consideration of this subject by others, especially from the perspective of their individual M&S communities.

This report presents a workable definition of validation and accreditation assessment referent that should be appropriate for all M&S varieties and applications.

This report stresses the importance of explicit and thorough referent specification, especially in regard to M&S intended use, information sources used for the referent and how they are combined, and how the referent is to be used in validation and accreditation assessments.

This report suggests ways that the referent should be described, and explores how M&S variety and application may impact referent description.

This report identifies three areas of important research that should be pursued to improve our capabilities for validation assessment referent identification and specification: how to consolidate information from various sources for use in a referent, how to specify dynamic referents, and how to document referents properly in M&S validation and accreditation.

8. References/Sources

23rd Army Science Conference, "Validation Of Medical Simulation Training," Session IA-05 (abstract only), December 2-5, 2002, available at http://www.asc2002.com/Abstracts_only/i/IA-05.pdf (accessed May 2004).

James H. Bigelow and Paul K. Davis, *Implications for Model Validation of Multi-Resolution Modeling (MRM) and Exploratory Analysis*, RAND Corporation Report DRR-3045-AF, April 2003.

G. E. P. Box, "Robustness is the Strategy of Scientific Model Building" in *Robustness in Statistics*, eds., R.L. Launer and G.N. Wilkinson, Academic Press, 1979, p. 202.

A. C. Calder et al, "A Case Study of Verifying and Validating an Astrophysical Code," in *V&V State of the Art: Proceedings of Foundations '02, a Workshop on Model and Simulation Verification and Validation for the 21st Century*, October 22-24, 2002, Laurel, MD. CD published by The Society for Modeling and Simulation, 2002.

Center for Integrative Multiscale Modeling and Simulation (CIMMS) Focused Workshop on Uncertainty in Engineering Design, May 23, 2002, Caltech, http://www.cimms.caltech.edu/workshops_dir/w-focused/abstracts/all_abstracts.pdf (accessed February 2004).

Nuno David, Jaime Simão Sichman, and Helder Coelho, "Towards an Emergence-Driven Software Process for Agent-Based Simulation," Third International Workshop on Multi-Agent Based Simulation (MABS'02, 15 July 2002), held at First International Joint Conference on Autonomous Agents and Multiagent Systems, 15 - 19 July 2002, Bologna – Italy. Available at http://www.lti.pcs.usp.br/SimCog/Doc/Pub/Papers/DavidFinalSpringerMABS02.pdf (accessed February 2004).

Department of Defense (DoD), *DoD Recommended Practices Guide (RPG) for Modeling and Simulation VV&A*, Millennium Edition (http://www.msiac.dmso.mil/vva/default.htm).

Roger Ghanem and Steve Wojtkiewicz, Uncertainty Quantification Short Course (February 9, 2003) at Society for Industrial and Applied Mathematics (SIAM) 2003 Conference on Computational Science and Engineering, http://www.siam.org/meetings/cse03/UQShortcourse.htm (accessed February 2004).

Daniel Girardot and René Jacquart, "A Proposed Evolution of Validation Definition," a T3 Paper in *V&V State of the Art: Proceedings of Foundations '02, a Workshop on Model and Simulation Verification and Validation for the 21st Century*, October 22-24, 2002, Laurel, MD. CD published by The Society for Modeling and Simulation, 2002.

David C. Gross, "Report of the Fidelity Implementation Study Group," *Proceedings of the Spring 1999 Simulation Interoperability Workshop (SIW)*, paper 99S-SIW-167A, 1999.

S. Y. Harmon, C. W. D. Hoffman, A. J. Gonzalez, R. Knauf, and V. B. Barr, "Validation of Human Behavior Representations," a Session B3 paper in *V&V State of the Art: Proceedings of Foundations '02, a Workshop on Model and Simulation Verification and Validation for the 21st Century*, October 22-24, 2002, Laurel, MD. CD published by The Society for Modeling and Simulation, 2002.

J. S. Hodges and J. A. Dewar, *Is It You or Your Model Talking? A Framework for Model Validation*, RAND Report, R-4114-AF/A/OSD, 1992.

Institute Electrical and Electronics Engineers (IEEE) Standard 1516.3, *Recommended Practice for High Level Architecture (HLA) Federation Development and Execution Process (FEDEP)*, 2003.

Regina Kunz and Andrew D. Oxman, "The Unpredictability Paradox: Review of Empirical Comparisons of Randomized and Non-randomized Clinical Trials," *British Medical Journal*, October 31, 1998, pp. 1185-1190. Available at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9794851&dopt=Abstract (accessed May 2004).

Federal Motor Carrier Safety (MCS) Administration, *Validation of Simulation Technology in the Training, Testing, and Licensing of Tractor-Trailer Drivers*, National Technical Information Service (NTIS), order number PB2000-100587, 2000. Also available from http://isweb.tasc.dot.gov, report number: FHWA-MC-99-060, and a short tech brief based on it is available at http://www.fmcsa.dot.gov/pdfs/tb00-007.pdf (accessed May 2004).

William L. Oberkampf, Sharon M. DeLand, Brian M. Rutherford, Kathleen V. Diegert, and Kenneth F. Alvin, *Estimation of Total Uncertainty in Modeling and Simulation*, Sandia National Laboratories Report SAND2000-0824, April 2000.

William L. Oberkampf, Timothy G. Trucano, Charles Hersch, "Verification, Validation, and Predictive Capability in Computational Engineering and Physics," Paper B1 in *V&V State of the Art: Proceedings of Foundations '02, a Workshop on Model and Simulation Verification and Validation for the 21st Century*, October 22-24, 2002, Laurel, MD. CD published by The Society for Modeling and Simulation, 2002.

Dale K. Pace and Jack Sheehan, "Subject Matter Expert (SME)/Peer Use in M&S V&V," Paper A4 in *V&V State of the Art: Proceedings of Foundations '02, a Workshop on Model and Simulation Verification and Validation for the 21st Century*, October 22-24, 2002, Laurel, MD. CD published by The Society for Modeling and Simulation, 2002.

Judea Pearl, *Simpson's Paradox: An Anatomy*, UCLA Cognitive Systems Laboratory, Technical Report (R-264), November 1999. Available at http://singapore.cs.ucla.edu/R264.pdf (accessed May 2004).

Patrick J. Roache, *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers (Albuquerque, NM), 1998. ISBN 0-913478-08-3.

Z. C. Roza, D. C. Gross, and S. Y. Harmon, "Report Out of the Fidelity Experimentation ISG," *Proceedings of the Spring 2000 Simulation Interoperability Workshop (SIW)*, paper 00S-SIW-151, 2000.

SAMO 2004, 4th International Conference on Sensitivity Analysis of Model Output, SAMO 2004,March 8-11, 2004, http://www.floodrisknet.org.uk/events/SAMO%202004 (accessed February 2004).

"Simulation Conceptual Model Development and Validation," Special Topic in *DoD M&S VV&A Recommended Practices Guide (RPG)*, Millennium Edition, 2000. Available at: http://vva.dmso.mil.

Star Logo, developed by MIT's Media Lab, http://education.mit.edu/starlogo/starterpage.html.

SimCity, originated in the late-1980s, has sold millions of copies, and has numerous websites related to it. The user can start the city and organize it, but the "sims" (its simulated citizens) and their behavior determine the way it evolves.

TNO Human Factors, *Project Example: Validation of the ANWB Driving Simulation*, January 20, 2004, Available from http://www.tm.tno.nl/product/res_to_19.html (accessed May 2004).

A. C. Turing, "Computing Machinery and Intelligence," *MIND: A Quarterly Review of Psychology and Philosophy*, Vol. LIX, No. 236, October 1950, pp. 433-460, available from http://www.abelard.org/turpap/turpap.htm (accessed May 2004).

Appendix A
Acronyms and Definitions

Acronyms and definitions are mixed alphabetically. Most of the definitions in this appendix were taken from the Glossary in the *DoD M&S VV&A Recommended Practices Guide (RPG).* In some cases, the *RPG* has more than one definition for a term, and the one used here is usually the most commonly used one, or at least consistent with the most common connotation for the term. Some definitions were those used by special topics in the *RPG.*

Accreditation The official certification that a model, simulation, or federation of models and simulations and its associated data are acceptable for use for a specific purpose.

Constructive Models and Simulations Models and simulations that involve simulated people with simulated systems. Real people make inputs to the simulation but are not involved in determining outputs of the simulation.

DoD Department of Defense

EuroSIW European Simulation Interoperability Workshop

Fidelity The degree to which a model or simulation reproduces the state and behavior of a real-world object or the perception of a real-world object, feature, condition, or chosen standard in a measurable or perceivable manner; a measure of the realism of a model or simulation; faithfulness. Fidelity should generally be described with respect to the measures, standards, or perceptions used in assessing or stating it.

GNP Gross National Product

HBR Human Behavior Representation

HWIL Hardware-in-the-Loop

IEEE Institute of Electrical and Electronics Engineers

Live Simulation A simulation involving real people operating real systems.

M&S This term has two connotations: 1) for Model/Models and/or Simulation/Simulations. The use of particular models and/or simulations, either statically or over time, to develop data as a basis for making managerial or technical decisions. This includes but is not limited to, emulators, prototypes, simulators, and stimulators. 2) for Modeling and/or Simulation. Addresses the general topic of modeling and simulation.

MCS Federal Motor Carrier Safety (MCS) Administration

Model A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process.

Referent *The referent is the best or most appropriate codified body of information available that describes characteristics and behavior of the reality represented in the simulation from the perspective of validation assessment for intended use of the simulation*. [proposed by this Report]

Simulation A method for implementing a model over time.

Simulation Conceptual Model The developer's description of what the model or simulation will represent, the assumptions limiting those representations, and other capabilities needed to satisfy the user's requirements. [From *RPG* special topic: A simulation conceptual model is a Developer's way of translating the requirements into a detailed design framework, from which the software that will make up the simulation can be built. A simulation conceptual model is the collection of information which describes a Developer's concept about the simulation and its pieces. That information consists of assumptions, algorithms, relationships (i.e., architecture), and data. Taken together, these items describe how the Developer understands what is to be represented by the simulation (e.g., entities, actions, tasks, processes, interactions) and how that representation will satisfy the requirements to which the simulation responds. A simulation conceptual model can be a primary mechanism for clear communication among simulation design and implementation personnel (e.g., systems analysts, system engineers, software designers, code developers, testers), Users, subject matter experts (SMEs) involved in simulation reviews, and verification, validation, and accreditation (VV&A) personnel.]

SISO Simulation Interoperability Standards Organization

SIW Simulation Interoperability Workshop

SME Subject Matter Expert.

Subject Matter Expert From *RPG* special topic: An individual who, by virtue of position, education, training, or experience, is expected to have greater-than-normal expertise or insight relative to a particular technical or operational discipline, system, or process, and who has been selected or appointed to participate in development, verification, validation, accreditation, or use of a model or simulation.

Validation The process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model.

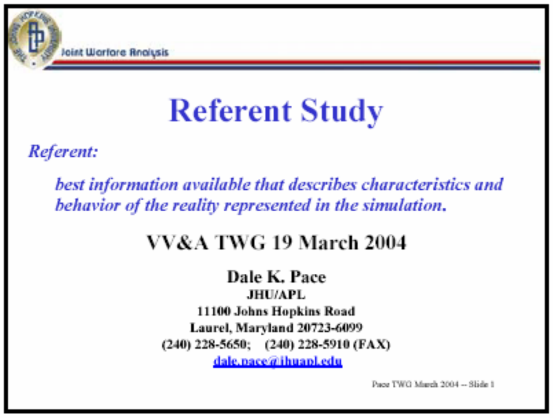
Verification The process of determining that a model implementation and its associated data accurately represent the developer's conceptual description and specifications.

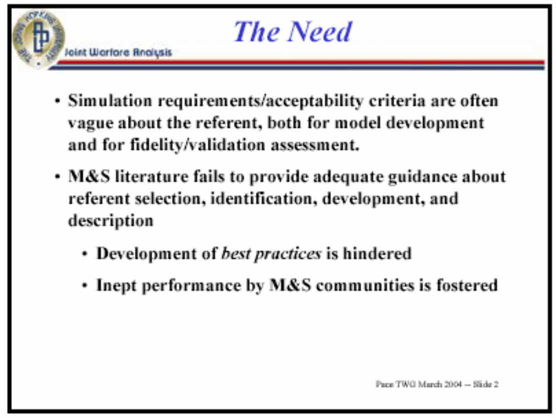
Virtual Simulation A simulation involving real people operating simulated systems.

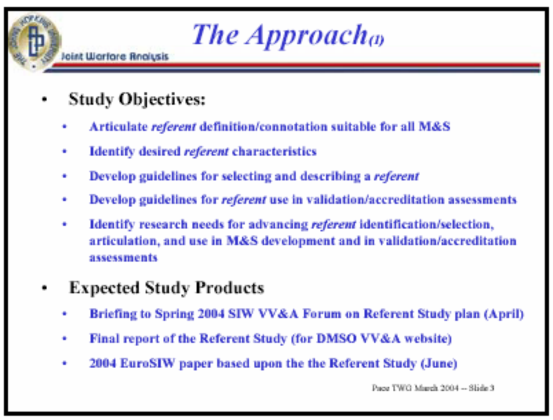
V&V Verification and Validation

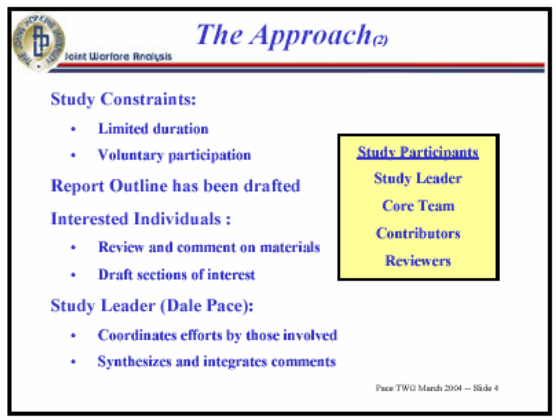
VV&A Verification, Validation, and Accreditation

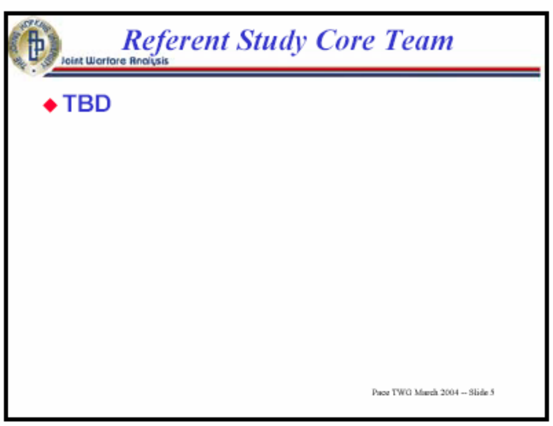
Appendix B
March Briefing to DoD VV&A TWG on Referent Study Plans

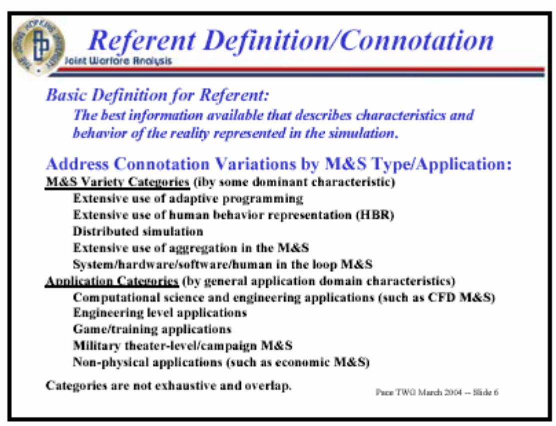












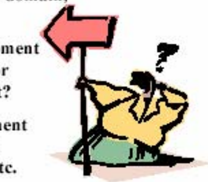
Joint Warfare Analysis

Selected Referent Issues

How should referent guidelines vary with M&S type or application domain, data availability, etc.?

How should the M&S development referent relate to the fidelity or validation assessment referent?

What is an acceptable assessment referent for a simulation with adaptive, emergent, chaotic, etc. behavior?



Pace TWG March 2004 -- Slide 7

Joint Warfare Analysis

The Opportunity

All who would like to comment upon materials developing in this study are welcome to do so (just send an email to Dale Pace indicating that).

- Materials will be sent to all interested –BUT there will be a short suspense date

All who wish to contribute to the study are welcome (just send an email to Dale Pace indicating that and specifying what area of the subject interests you most and how much time you would expect to be able to devote to the study in April & May).

Dale.Pace@jhuapl.edu

Pace TWG March 2004 -- Slide 8

Joint Warfare Analysis

Questions & Comments

Pace TWG March 2004 -- Slide 9

Joint Warfare Analysis

Back-up Slides

Pace TWG March 2004 -- Slide 10

Joint Warfare Analysis

Report Outline

- Front Matter (Title Page, Table of Contents, Acknowledgements, etc.)
- Executive Summary
- Introduction
  - Background (lack of referent discussion in M&S/VV&A literature, and general failure to be very explicit about referents in M&S assessments)
  - Objectives and Scope of Study
  - Report Organization
- Referent Definition and Connotation
- How to Identify Possible Referents and Select an Appropriate Referent
- How to Specify/Describe the Referent
- Referent Use in M&S Development and in Validation and Accreditation Assessments
  - With data-rich referents,
  - With data-poor referents, and
  - With poorly defined/identified or otherwise inadequate referents
- Referent Research Needs
- Conclusion
- References/Sources
- Appendices (as needed)

Pace TWG March 2004 -- Slide 11

Appendix C
April Briefing to SIW VV&A Forum & PDG on Referent Study Plans

This briefing will be distributed via the SIW VV&A reflector.

Joint Warfare Analysis

Referent Study

Referent:

The referent is the best or most appropriate codified body of information available that describes characteristics and behavior of the reality represented in the simulation from the perspective of validation assessment for intended use of the simulation.

SIW (April 2004)

Dale K. Pace
JHU/APL
11100 Johns Hopkins Road
Laurel, Maryland 20723-6099
(240) 228-5650; (240) 228-5910 (FAX)
dale.pace@jhuapl.edu

Joint Warfare Analysis

The Need

- Simulation requirements/acceptability criteria are often vague about the referent, both for model development and for fidelity/validation assessment.
- M&S literature fails to provide adequate guidance about referent selection, identification, development, and description
  - Development of *best practices* is hindered
  - Inept performance by M&S communities is fostered

Joint Warfare Analysis

The Approach(1)

- Study Objectives:
  - Articulate *referent* definition/connotation suitable for all M&S
  - Identify desired *referent* characteristics
  - Develop guidelines for selecting and describing a *referent*
  - Develop guidelines for *referent* use in validation/accreditation assessments
  - Identify research needs for advancing *referent* identification/selection, articulation, and use in M&S development and in validation/accreditation assessments
- Expected Study Products
  - Briefing to Spring 2004 SIW VV&A Forum on Referent Study plan (April)
  - Final report of the Referent Study (for DMSO VV&A website)
  - 2004 EuroSIW paper based upon the Referent Study (June)

Joint Warfare Analysis

The Approach(2)

Study Constraints:

- Limited duration
- Voluntary participation

Report Outline has been drafted

Interested Individuals :

- Review and comment on materials
- Draft sections of interest

Study Leader (Dale Pace):

- Coordinates efforts by those involved
- Synthesizes and integrates comments

Study Participants
Study Leader
Core Team
Contributors
Reviewers

Joint Warfare Analysis

Referent Study Core Team

- Dale K.Pace (Johns Hopkins University Applied Physics Laboratory) – Study Lead
- Mark Dumble (UK)
- Hanae Hara (SPAWAR)
- Scott Harmon (Zetetix)
- Sam Johnson (USAF M&S)
- Mike Leite (DMSO, SAIC)
- Mike Metz (IMC)
- Bob Senko (DMSO Consultant)
- Susan Solick (TRAC)
- Marcy Stutzman (Northrop Grumman)
- Scott Weidman (National Academy of Sciences)
- Simon M. Youngblood (DMSO VV&A Technical Director)

Joint Warfare Analysis

Referent Definition

Initial Definition for Referent Study:

The referent is the best or most appropriate codified body of information available that describes characteristics and behavior of the reality represented in the simulation from the perspective of validation assessment for intended use of the simulation.

Issue: Often "best info" is too expensive; many use "acceptable info"

Challenge: How to determine "acceptable info"

We hope the study will provide useful guidance in this area.

Referent Connotation

Joint Warfare Analysis

Address Connotation Variations by M&S Type/Application:

M&S Variety Categories (by some dominant characteristic)
- Extensive use of adaptive programming
- Extensive use of human behavior representation (HBR)
- Distributed simulation
- Extensive use of aggregation in the M&S
- System/hardware/software/human in the loop M&S

Application Categories (by general application domain characteristics)
- Computational science and engineering applications (such as CFD M&S)
- Engineering level applications
- Game/training applications
- Military theater-level/campaign M&S
- Non-physical applications (such as economic M&S)

Categories are not exhaustive and overlap.

Pace SIW April 2004 -- Slide 7

Examples of Referent Issues

Joint Warfare Analysis

How should referent guidelines vary with M&S type or application domain, data availability, etc.?

What is an acceptable assessment referent for a simulation with adaptive, emergent, chaotic, etc. behavior?

Pace SIW April 2004 -- Slide 8

The Opportunity

Joint Warfare Analysis

Any who would like to review & comment upon materials developed for this study are welcome to do so (just send an email to Dale Pace indicating that).

- Materials will be sent to all interested

  BUT the suspense date will be short (about 10 days!)

Dale.Pace@jhuapl.edu

Any who wishes to contribute to a specific area of the study should contact Dale Pace to see how that can be arranged. Such contributions would have to be provided within the next 2-3 weeks.

Pace SIW April 2004 -- Slide 9

Report Initial Outline

Joint Warfare Analysis

- Front Matter (Title Page, Table of Contents, Acknowledgements, etc.)
- Executive Summary
- Introduction
  - Background (lack of referent discussion in M&S/VV&A literature, and general failure to be very explicit about referents in M&S assessments)
  - Objectives and Scope of Study
  - Report Organization
- Referent Definition and Connotation
- How to Identify Possible Referents and Select an Appropriate Referent
- How to Specify/Describe the Referent
- Referent Use in M&S Development and in Validation and Accreditation Assessments
  - With data-rich referents,
  - With data-poor referents, and
  - With poorly defined/identified or otherwise inadequate referents
- Referent Research Needs
- Conclusion
- References/Sources
- Appendices (as needed)

Pace SIW April 2004 -- Slide 10

Questions & Comments

Joint Warfare Analysis

Pace SIW April 2004 -- Slide 11